Stuart, Kari, Kozan, Erhan, Sinnott, Michael, & Collier, James (2010) ***An innovative robust reactive surgery assignment model***. ASOR Bulletin, 29(3), pp. 48-58.

An Innovative Robust Reactive Surgery Assignment Model

Kari Stuart*, Erhan Kozan*, Michael Sinnott and James Collier****

[*] Discipline of Mathematical Sciences, Queensland University of Technology

GPO Box 2434, Brisbane Qld 4000, Australia

k.stuart,e.kozan@qut.edu.au

[**] Princess Alexandra Hospital Emergency Department

Brisbane Qld 4102 Australia

Michael_Sinnott, James_Collier@health.qld.gov.au

Abstract.

Online scheduling in the Operating Theatre Department is a dynamic process that deals with both elective and emergency patients. Each business day begins with an elective schedule determined in advance based on a mastery surgery schedule. Throughout the course of the day however, disruptions to this baseline schedule occur due to variations in treatment time, emergency arrivals, equipment failure and resource unavailability. An innovative robust reactive surgery assignment model is developed for the operating theatre department. Following the completion of each surgery, the schedule is re-solved taking into account any disruptions in order to minimise cancellations of pre-planned patients and maximise throughput of emergency cases. The single theatre case is solved and future work on the computationally more complex multiple theatre case under resource constraints is discussed.

Key Words: Robust Reactive Assignment, Online, Operating Theatre.

Introduction

The Operating Theatre (OT) department is a dynamic environment consisting of pre-scheduled elective patients and unpredictable emergency cases. Nearly every department within a hospital schedules patients for the Operating Theatre (OT) and most wards receive patients from the OT following post-operative recovery. Because of the interrelationships between units, disruptions and cancellations within the OT can have a flow on effect to the rest of the hospital. This often results in dissatisfied patients, nurses and doctors, escalating waiting lists, inefficient resource usage and undesirable waiting times. Improving the overall responsiveness to emergency patients by solving the disruption management and re-scheduling problem is used in this paper to improve the efficiency of the OT.

To date, most literature has focused on the scheduling of elective patients, which has been

† : Corresponding Author

approached in many ways. Dexter and Traub (2000) applied decision theory to the process of sequencing patients. Pham and Klinkert (2008) treated surgical case scheduling as a generalised job shop scheduling problem. Galvin (1997) applied the cutting stock problem to the surgical scheduling problem. Persson and Persson (2006) used optimisation modelling to synchronise allocation of different resources for operating room planning. Sier et al., (1997) developed a tool for scheduling operations that considered bed availability, efficient theatre utilisation, minimising schedule deviations and emergency arrivals.

Although many good theoretical approaches exist for elective scheduling, many of these fail in the online setting due to disturbances including variability in surgical durations and the arrival of emergency patients. To address this issue, schedulers are moving away from deterministic and stochastic optimisation scheduling models towards robust scheduling models (Daniels and Kouvelis, 1995). Robust schedules are generated such that the schedule performance remains high even in the presence of online disruptions (Leon et al., 1994). This may be achieved by the inclusion of 'buffers' that absorb variations in treatment times that occur during project execution. Hans et al., (2008) introduced robust scheduling to the OT by assigning elective surgeries and planned slack time to the operating room days to prevent overtime. The planned slack time on each operating room-day is based on the expected variance of the surgical durations planned on that day. The effect of this is to create a schedule whose performance is relatively insensitive to the potential realisations of the task parameters (Daniels and Kouvelis, 1995).

Robust scheduling is an example of a preventive or proactive scheduling approach that serves as a baseline schedule for online production scheduling. Effective preventive schedules are important since they form the basis for resource commitment decisions. When used in conjunction with reactive scheduling models, they improve the performance of online scheduling (Li and Ierapetritou, 2008).

While robust schedules address quality robustness, they do not address solution robustness (Van de Vonder et al., 2005). Disruptions during project execution may cause deviations from a predictive schedule and even make it infeasible. Solution robustness is addressed by reactive scheduling, which is used to repair the baseline schedule following activity disruptions, by including changes whilst minimising disruptions from the original schedule. Two types of reactive scheduling include the repair of the existing schedule and full scheduling of tasks after a disruptive event i.e. re-scheduling (Li and Ierapetritou, 2008, Sabuncuoglu and Bayiz, 2000). To the authors' knowledge, there is no literature to date dealing with reactive scheduling of elective and emergency patients for the operating theatre.
For operating theatre scheduling, two types of disruptions are defined; a theatre (machine) disruption and patient (job) disruption. Theatre disruptions occur when a theatre becomes

unavailable for some period of time. Examples include equipment failure or staff shortage/unavailability or the arrival of a high priority emergency patient that requires the use of an elective theatre. This type of disruption results in the patients that were initially scheduled for that theatre (and have not yet been treated) to be delayed and the schedule must be updated to take into account such changes. Patient disruptions on the other hand occur when treatment times are less than or greater than the assigned surgery time. If surgery duration is shorter than expected, this generally means the schedule can be moved forwards without much alteration. Also, there may be time left at the end of the schedule for adding on emergency cases, or the additional time may be spent on a patient that exceeds its expected duration. If patients exceed their expected duration however, this can cause delays in surgery start time for remaining patients or may even necessitate cancellation of remaining surgeries to prevent overtime of the theatre.

Introduced is an innovative robust reactive assignment model (RRAM) for dealing with online disruptions in order to minimise cancellations of pre-planned patients and maximise throughput of emergency cases. This is achieved by keeping track of the immediately preceding schedule, before the completion of an operation. The original baseline schedule is developed using a robust scheduling approach that assumes surgical durations are lognormally distributed. This schedule provides the list of patients to be assigned to an operating theatre.

The Robust Reactive Assignment Model

The robust reactive assignment model (RRAM) is an assignment model that aims to minimise cancellations of the already assigned patients and maximise throughput of emergency cases following disruptions in the online environment. Disruptions include but are not necessarily limited to variations in a patient's estimated treatment time, the arrival of an emergency patient, equipment failure and resource unavailability (staffing or equipment). Regardless of the type of disruption, these lead to either early or late start times for the remaining patients. For example a patient that is completed early or an unexpected cancellation of an earlier patient may lead to an early start for the next patient. Conversely, a late finish or a delay in resource availability may lead to a late start for the subsequent patient/s or result in cancellations.

At the completion of each patient's surgery the RRAM is implemented. The model takes into account deviations between the actual duration of the surgery and the amount of time assigned to the procedure and any emergency arrivals during the course of the schedule. For example, if the procedure takes longer than expected, then there is less time available for the remaining patients on the schedule list. This is reflected in the amount of time remaining in the theatre, and some patients may need to be postponed if there is insufficient time to schedule all patients on the current list.

The objective of the reactive schedule is to minimise the costs of the new schedule brought on by cancellation of patients and offset by the inclusion of additional cases. The cost of cancelling (or profit earned by assigning) a patient depends on their priority level and whether or not they are an emergency or elective patient. Patients with higher priority receive a higher penalty for cancellation (or profit for inclusion).

Notations

i : Specialties considered within a surgical category, $i \in \{1,..,I\}$

j : Priority, $j \in \{1,..,J\}$

k : Type of patient (elective or emergency), $k \in \{1,..,K\}$

d_i : Expected surgical duration of specialty i

s_i^2 : Expected surgical duration volatility of specialty i

μ_i, σ_i : Lognormal distribution parameters for specialty i

M : The sum of the expected surgical durations assigned

V : The sum of the expected surgical durations assigned

T : Time remaining in the theatre

E_{ijk} : The number of available patients of specialty i, priority j and type k

X_{ijk} : The number of initially assigned patients of specialty i, priority j and type k

$X_{ijk}^{'}$: The number of assigned patients of specialty i, priority j and type k following the reschedule.

C_{jk} : Cost/benefit of patients of priority j and type k

Objective function

The objective of the reactive schedule is to minimise the costs of the new schedule. These costs include the cost of cancelling a patient, the profit (or negative cost) earned for adding a patient and a penalty for deviating from the original schedule. The penalty incurred for cancelling a patient (or profit earned for adding one) depends on their priority level *j* and type *k*. This is represented by $C_{jk}\left(X_{ijk} - X_{ijk}^{'}\right)$. If $\left(X_{ijk} - X_{ijk}^{'}\right)$ is positive, then at least one patient has been cancelled and a penalty, C_{jk} is incurred for each cancellation.

If $\left(X_{ijk} - X_{ijk}^{'} \right)$ is negative, then at least one patient of specialty *i*, priority *j* and type *k* has been added to the schedule and a profit C_{jk} per patient is subtracted from the objective function. The second aspect of the objective function, $\left| X_{ijk} - X_{ijk}^{'} \right|$, prevents the addition of patients at the expense of cancelling another. For example, one patient with penalty of 10 units could be cancelled and replaced with two patients, each with a profit of 5 units. In practice, this would generally be considered both impractical and destructive to the efficiency of the schedule.

$$\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} C_{jk} \left(X_{ijk} - X_{ijk}^{'} \right) + \left| X_{ijk} - X_{ijk}^{'} \right| \quad \textbf{(1)}$$

Constraints

The number of patients assigned in the new schedule cannot exceed the number available. This allows for a cancelled elective patient to be re-assigned at a later event.

$$X_{ijk}^{'} \leq E_{ijk} \quad \textbf{(2)}$$

For model simplification surgical duration mean and volatility of mean estimates are assumed independent on priority level *j* and whether the patient is an emergency or elective. Historical patient data was analysed and surgical durations are modelled with a lognormal distribution. The expected surgical duration, d_i and variance, s_i^2 of specialty *i*, are respectively given by

$$d_i = e^{\mu_i + \frac{\sigma_i^2}{2}} \quad \textbf{(3)}$$

$$s_i^2 = \left(e^{\sigma_i^2} - 1 \right) e^{2\mu_i + \sigma_i^2} \quad (4)$$

where μ_i and σ_i are lognormal random variable parameters determined by analysis of historical data. The sum of the expected durations and the variance of the patients assigned to the theatre are given respectively by

$$M = \sum_{i=1}^{I} \left(\left(\sum_{j=1}^{J} \sum_{k=1}^{K} X_{ijk}^{'} \right) e^{\mu_i} \sqrt{e^{\sigma_i^2}} \right) \quad \textbf{(5)}$$

$$V = \sum_{i=1}^{I} \left(\left(\sum_{j=1}^{J} \sum_{k=1}^{K} X_{ijk}^{'} \right) \left(e^{\sigma_i^2} - 1 \right) e^{2\mu_i + \sigma_i^2} \right) \quad \textbf{(6)}$$

The amount of time that is planned for the patients assigned to the theatre depends on the level of accuracy desired by the decision maker. The level of accuracy used for the model is 15.87%, i.e. the probability that surgeries run overtime is less than 15.87%. The amount of time that is planned for the surgeries assigned to the theatre, based on this level of accuracy is given by

$$\left(\frac{M^2}{\sqrt{V+M^2}}\right)e^{\sqrt{\ln\left(\frac{V+M^2}{M^2}\right)}} \tag{7}$$

Equation 8 ensures the time available in the theatre is sufficient to complete all the assigned surgeries.

$$T \geq \left(\frac{M^2}{\sqrt{V+M^2}}\right)e^{\sqrt{\ln\left(\frac{V+M^2}{M^2}\right)}} \tag{8}$$

The number of patients assigned to each theatre is a positive integer

$$X'_{ijk} = \text{integer},\ \forall i, j, k \tag{9}$$

Solution approach and results

The RRAM is a combinatorial optimisation problem and is a variation of a bounded knapsack problem in which the patients are the objects to be assigned to a theatre (knapsack) with limited capacity. The number of patients that can be assigned of a particular specialty, priority and type is bounded by variable E_{ijk}. Each patient takes up a portion of the limited capacity and has a benefit, C_{jk} associated with its assignment. However, in this case, the amount of time assigned for a given patient specialty is not fixed, but depends on the existing assignment of patients (see Equations 5 – 7).

The generalised bounded knapsack problem maximises the benefit associated with assigning the packed items. In our model, this is analogous to minimising $\left(-\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} C_{jk} X'_{ijk}\right)$. In order to minimise the costs of our objective, one must also solve the bounded knapsack problem, however, we have the added consideration of minimising changes from the original assignment of patients. Therefore our problem is just as computationally difficult as the bounded knapsack problem.

Knapsack problems have received much attention in the literature and are known to be NP hard in the ordinary sense and may be solved in pseudo-polynomial time. An algorithm that runs in pseudo-polynomial time has a running time that is polynomial in the numeric value of the input. Thus, as the number of patients increases, so does the computation time. However, in the case of the single operating theatre, the number of patients is relatively small and therefore can be solved without the use of metaheuristics.

Various exact solution methods exist including dynamic programming, branch and bound and exhaustive methods. Dynamic programming requires the problem to be broken up into stages that can be solved individually. Due to the robust calculations of processing time, dynamic programming cannot be applied to the RRAM. Branch and bound, commercial software or an enumerative approach could be used for the RRAM. The benefit of an enumerative algorithm over commercial software and branch and bound, is that the code can easily be adapted with changes in the problem structure. For example, if the problem is expanded to the multiple theatre case, the problem becomes analogous to a multiple knapsack problem or bin packing problem. For this reason, an enumerative algorithm is used and implemented using Visual Basic.

Implementing the model in Visual Basic has the additional benefit of a user-friendly interface, designed for use within the practical setting. The model is run after the completion of each operation. The user is prompted for information on the duration of the completed surgery, the type of surgery (elective or emergency), and the specialty and priority of the patient treated. In addition, information on emergency patient arrivals can be added at any time.

The enumerative algorithm is used to search for the best schedule based on the information supplied. The objective is to minimise changes from the original schedule whilst also searching for the schedule that produces the best objective value. This is important because it is not practical from a resource viewpoint to make large changes to the original schedule.

The RRAM was applied to a surgical care unit (SCU) (or day surgery unit). Day surgery patients generally arrive and/or are released on the day of their surgery. Using the SCU is a satisfactory representation of the whole OT department because both elective and emergency patients are treated there. The benefit to using the SCU is the reduction in problem size and hence calculations. The results obtained for the SCU can be extended and applied to the entire OT department.

The RRAM was tested by simulating the implementation of 100 offline robust schedules for a day surgery unit. Historical data was analysed to determine appropriate statistical distributions to generate surgical durations and also to determine the arrival pattern of emergency patients. Since the SCU was selected as a sample for the whole problem, specialties that are only treated in the SCU were sought after, to get a closer representation of the real life problem. If a theatre only treats a particular group (or category) of patients then the total capacity is dedicated to those patients. In other words, there is no need for adjusting total capacity available to those patients. If a number of different categories are treated in a theatre, then either all categories need to be considered, or the capacity must be adjusted for those selected. Of the categories treated in the SCU, the Ophthalmology patients are almost exclusively treated in two of the SCU theatres. Of the categories treated in these two theatres, the majority are also Ophthalmology. Most other categories (treated in the SCU) are also spread across the non-day surgery theatres. For this reason, the model is applied to the Ophthalmology patients and it is assumed that all of the available capacity is dedicated to those patients. This assumption may be changed according to a scheduler's requirements. It is important to note that the developed models may also be adjusted to incorporate any patient type and any theatre.

Historical data for the Ophthalmology patients was analysed for surgical duration estimates. The times available in the data provided were 'time in suite', 'in anaesthesia', 'in OR' and 'Out OR'. These times indicate the time the patient enters the operating theatre suite (which is composed of the operating rooms and their anaesthesia workrooms, day surgery unit, ICU, PACU etc), the time the patients enters anaesthesia and the times of entering and exiting the operating room respectively. The time of anaesthesia may be calculated as the difference between 'in OR' and 'in anaesthesia'. Likewise, the time in the OR is the difference between 'Out OR' and 'in OR'. These varied according to the type of surgery being performed. The total processing time of a patient was assumed to include anaesthesia preparation time, however this assumption can easily be changed if desired.

The Ophthalmology category was broken down into specialties for the calculation of surgical duration estimates. Six surgical specialties were determined based on the type of surgery performed. A histogram of actual data suggested a possible lognormal distribution for each of the surgical specialties. Goodness of fit tests supported the hypothesis that surgical durations may be described with the lognormal distribution. An example plot of the fitted distribution against actual data for the first specialty is provided in Figure 1.

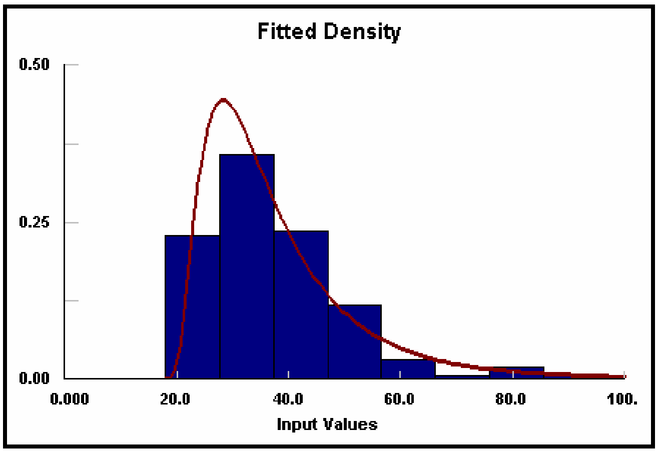

Figure 1. Fitted distribution versus actual data for specialty 1

The lognormal distribution is a continuous distribution, based on the normal distribution. It is used to describe many applications including physicians' consultation time, lifetime distributions, the long-term return rate on a stock investment and weight and blood pressure of humans. It has also been used in the literature for describing surgical durations (Jebali et al., 2006, Strum et al., 2000).

Parameters for the lognormal approximations were determined for each of the surgical specialties based on the data analysis and are presented in Table 1. The Lognormal distribution is described with 3 parameters, i.e. the mean of the included normal μ, the standard deviation of the included normal σ and the minimum value or location parameter γ. For data analysis, surgical durations were given in minutes.

Table 1. Data for surgical specialties

Specialty	Lognormal		
	γ	μ	σ^2
1	18	2.78	0.674
2	20	3.38	0.561
3	16	3.42	0.779
4	12	3.89	0.777
5	55	4.0	0.79
6	29	3.82	0.452

The elective patients for these test schedules were generated randomly, and assigned to the theatres using a multiple knapsack approach. Procedure durations, drawn from the appropriate lognormal distributions that were fitted from historical data, were generated for each patient. Emergency patients were randomly generated using an exponential distribution with an average inter-arrival time of 225 minutes (based on historical data). The results of the test cases are given in Table 2. The performance measures presented are the number of elective patients originally scheduled that were cancelled, the number of emergency cases performed, the number of electives cancelled but later re-scheduled and the number of emergencies added to the schedule that had to later be cancelled.

Table 2. Results of reactive schedules

Results	TOTAL
Number Electives Initially Assigned	741
Total Patients Completed	793
Number Emergencies Completed	178
Number Electives Cancelled	126
Re-scheduled Electives	71
Cancelled Emergencies	34

Results indicate that across the 100 examples there were a total of 126 elective cancellations and the RRAM was able to re-schedule 71 electives. The ability to re-schedule patients, when an already completed patient uses less time than it was allocated, illustrates the benefit of the robustness built into the model. In addition, by allowing a calculated amount of extra time for each surgery based on a percentage determined by the decision maker, the number of cancellations is kept low and allows for additional emergency patients to be seen. In this case, the model saw 178 additional emergencies performed across the 100 test cases. This ability to schedule additional patients ensures theatre capacity is used efficiently rather than being left unused. In 34 instances, emergency cases that had been tentatively assigned to a schedule had to be cancelled due to lack of time. Allowing for their cancellation also helps to maintain the efficiency of the theatre utilisation by preventing overruns. For the 100 test cases, only 16% resulted in over-run theatres.

In addition to measuring performance indicators, changes in the schedule may be presented in Gantt charts. Figure 2 illustrates the Gantt chart for one of the schedules. In the initial schedule, 9 elective patients are assigned to the theatre, each of which is illustrated with a different colour. After each surgery is completed, the RRAM is run and an updated schedule is generated. Following the late completion of the first patient (in yellow), the RRAM shifts the original schedule to the right and post-pones one elective patient (patient 8, in green) and fills the remaining capacity with an emergency patient (patient 10, in orange). The next seven procedures (patients 2 – 7 and 9) finish either early or on time. No more emergency patients are added to the schedule and patient 8 is not re-scheduled. After each of these procedures, the RRAM is implemented and there is either a 'left' shift in the schedule (for an early completion) or the schedule is unchanged (for on-time completions).
The final patient (patient 10, in orange) requires an additional 5 minutes of surgery time. It is evident from the Gantt chart that this particular schedule completes 'on-time' meaning that

it does not exceed the capacity of 480 minutes.

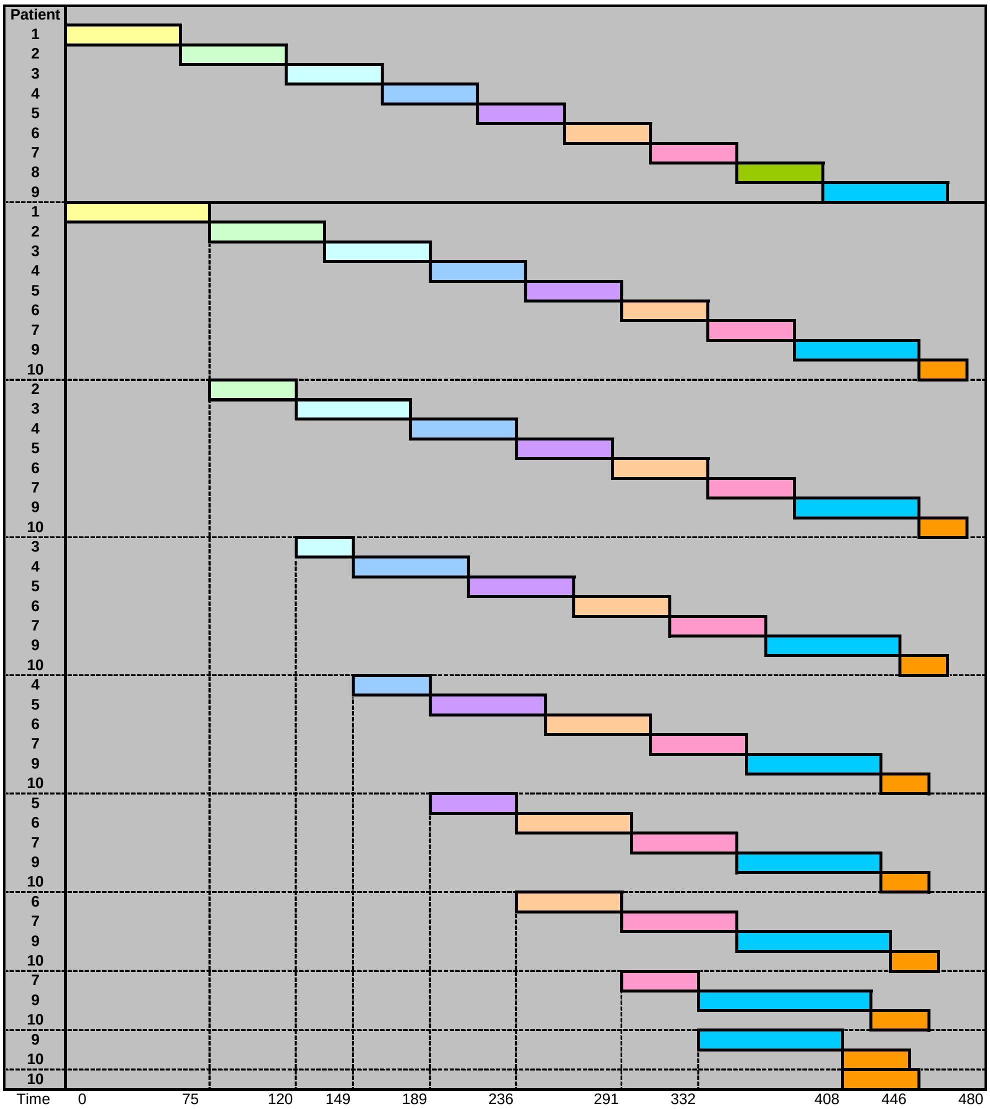

Figure 2. Gantt Chart illustrating results for one schedule

Conclusions

An innovative online assignment model for a single Operating Theatre is developed and solved. The model is run in real-time following the completion of each operation and minimises cancellations whilst also allowing for additional scheduling of emergency cases (time permitting), which may arise during the schedule's implementation. The problem is NP hard in the ordinary sense and hence an exact solution approach was used. The model was developed and implemented using Visual Basic.

Results for the RRAM showed it was capable of adapting appropriately to disruptions in the online environment by delaying, rescheduling or adding additional surgeries according to the available operating time capacity. The ability to re-schedule patients, when an already completed surgery used less time than it was allocated, illustrated the benefit of the robustness built into the model. In addition, by allowing a calculated amount of extra time for each surgery based on a percentage determined by the decision maker, the number of cancellations was kept low and additional emergency patients could be treated. This ability to schedule additional patients ensured theatre capacity is used efficiently rather than being left unused. Allowing for the cancellation of emergencies also helped to maintain the efficiency of the theatre utilisation by preventing overruns.

One limitation of this research is that it only considers the assignment of patients to a single operating theatre and does not consider patient sequence. Future work of the authors includes the adaptation of the model to include patient sequence and also to extend the problem to include multiple theatres.

As mentioned earlier in the paper, the objective of the model is to minimise the costs of the new schedule by minimising the number of elective cancellations and maximising the number of electives that may be added to the schedule. Because the model is re-run after each patient's completion it does not keep track of the preceding objective function values. The resulting schedule may be different if the objective function values were carried through for each schedule. For example, the number of elective cancellations may possibly be reduced with an accompanied decrease in emergencies added to the schedule. This could also form a future topic of research. Other changes to the schedule results could be induced by changes in the scheduler's objectives and should be investigated. For example, if an elective is cancelled relatively towards the beginning of a schedule, then the user may opt not to fill in the remaining capacity with available emergencies at that point in time and wait until later in the schedule.

REFERENCES

[1] R. L. Daniels and P. Kouvelis, "Robust scheduling to hedge against processing time uncertainty in single-stage production," *Management Science*, vol 41, pp.363-376, 1995.

[2] F. Dexter and R. D. Traub, "Sequencing cases in the operating room: predicting whether one surgical case will last longer than another," *Anesthesia & Analgesia*, vol 90, pp.975-979, 2000.

[3] W. Galvin, "Hospital theatre scheduling and cutting-stock problem," *ASOR*

Bulletin, vol 16, pp.3-12, 1997.

[4] E. Hans, G. Wullink, M. Van Houdenhoven and G. Kazemier, "Robust surgery loading," *European Journal of Operational Research*, vol 185, pp.1038-1050, 2008.

[5] A. Jebali, A. B. Hadj-Alouane and P. Ladet, "Operating rooms scheduling," *International Journal of Production Economics*, vol 99, pp.52-62, 2006.

[6] V. J. Leon, S. D. Wu and R. H. Storer, "Robustness measures and robust scheduling for job shops," *IIE Transactions*, vol 26, pp.32-43, 1994.

[7] Z. Li and M. G. Ierapetritou, "Process scheduling under uncertainty: Review and challenges," *Computers and Chemical Engineering*, vol 32, pp.715-727, 2008.

[8] M. Persson and J. A. Persson, "Optimization modeling of hospital operating room planning using a logistic perspective," *Proc. of 2006 Annual Conference of OR Applied to Health Services*, 2006.

[9] D.-N. Pham and A. Klinkert, "Surgical case scheduling as a generalized job shop scheduling problem," *European Journal of Operations Research*, vol 185, pp.1011-1025, 2008.

[10] I. Sabuncuoglu and M. Bayiz, "Analysis of reactive scheduling problems in a job shop environment," *European Journal of Operational Research*, vol 126, pp.567-586, 2000.

[11] D. Sier, P. Tobin and C. McGurk, "Scheduling surgical procedures," *Journal of the Operational Research Society*, vol 48, pp.884-891, 1997.

[12] D. P. Strum, J. H. May and L. G. Vargas, "Modeling the uncertainty of surgical procedure times: comparison of lognormal and normal models," *Anesthesiology*, vol 92, pp.1160-1167, 2000.

[13] S. Van de Vonder, E. Demeulemeester, W. Herroelen and R. Leus, "The use of buffers in project management: the trade-off between stability and makespan," *International Journal of production economics*, vol 97, pp.227-240, 2005.